Sparse Phase Retrieval: Uniqueness Guarantees and Recovery Algorithms

Kishore Jaganathan Samet Oymak Babak Hassibi*

Department of Electrical Engineering,
Caltech, Pasadena, CA - 91125.

kishore@caltech.edu, soymak@caltech.edu, hassibi@caltech.edu

Abstract

The problem of signal recovery from its Fourier transform magnitude, or equivalently, autocorrelation, is of paramount importance in various fields of engineering and has been around for over 100 years. In order to achieve this, additional structure information about the signal is necessary. In this work, we first provide simple and general conditions, which when satisfied, allow unique recovery almost surely. In particular, we focus our attention on sparse signals and show that most $O(n)$-sparse signals, i.e., signals with $O(n)$ non-zero components, have distinct Fourier transform magnitudes (up to time-shift, time-reversal and global sign). Our results are a *significant* improvement over the existing identifiability results, which provide such guarantees for only $O(n^{1/4})$-sparse signals.

Then, we exploit the sparse nature of the signals and develop a Two-stage Sparse Phase Retrieval algorithm (TSPR), which involves: (i) identifying the support, i.e., the locations of the non-zero components, of the signal (ii) identifying the signal values using the support knowledge. We show that the proposed algorithm can *provably* recover $O(n^{1/2-\epsilon})$-sparse signals (up to time-shift, time-reversal and global sign) with arbitrarily high probability in quadratic-time. To the best of our knowledge, state of the art phase retrieval algorithms provide such recovery guarantees for only $O(n^{1/4})$-sparse signals. Numerical experiments complement our theoretical analysis and verify the effectiveness of the proposed algorithms.

1 Introduction

In many physical measurement systems, the power spectral density of the signal, i.e., the magnitude square of the Fourier transform, is the measurable quantity. The phase information of the Fourier

*This work was supported in part by the National Science Foundation under grants CCF-0729203, CNS-0932428 and CIF-1018927, by the Office of Naval Research under the MURI grant N00014-08-1-0747, and by a grant from Qualcomm Inc.

transform is completely lost, because of which signal recovery is difficult. Recovering a signal from its Fourier transform magnitude is known as phase retrieval [1, 2]. This recovery problem is one with a rich history and occurs in many areas of engineering and applied physics, including X-ray crystallography [3], astronomical imaging [4], speech processing [5], optics [6], computational biology [7], blind channel estimation [41] and so on.

Suppose $\mathbf{x} = (x_0, x_1, ..., x_{n-1})$ is a real-valued discrete-time signal of length n. Let $\mathbf{y} = (y_0, y_1, ...y_{n-1})$ be its Fourier transform, i.e.,

$$\mathbf{y} = \mathbf{F}\mathbf{x} \tag{1}$$

where $\mathbf{F}$ is the DFT matrix. The phase retrieval problem can be mathematically stated as

$$\begin{aligned} &\text{find} && \mathbf{x} \\ &\text{subject to} && |\mathbf{y}| = |\mathbf{F}\mathbf{x}| \end{aligned} \tag{2}$$

Since magnitude square of the Fourier transform and the autocorrelation function are Fourier pairs, the phase retrieval problem can be equivalently stated as recovering a signal from its autocorrelation, i.e.,

$$\begin{aligned} &\text{find} && \mathbf{x} \\ &\text{subject to} && \mathbf{a} = \mathbf{x} \star \tilde{\mathbf{x}} \end{aligned} \tag{3}$$

where $\mathbf{a} = (a_0, a_1, ..., a_{n-1})$ is the autocorrelation of $\mathbf{x}$ and $\tilde{\mathbf{x}}$ is the time-reversed version of $\mathbf{x}$[§].

Observe that the operations of time-shift, time-reversal and global sign-change on the signal do not affect the autocorrelation (and hence the Fourier transform magnitude), because of which there are trivial ambiguities. These signals are considered equivalent, and in all the applications it is considered good enough if any equivalent signal is recovered. For example, in astronomy where the objective is to know about the stars in the sky, or in X-ray crystallography where the objective is to know about the atomic and molecular structure of a crystal, this ambiguity does not matter[*] .

Suppose $X(z)$ and $A(z)$ are the z-transforms of the signal $\mathbf{x}$ and its autocorrelation $\mathbf{a}$ respectively. We have

$$A(z) = X(z)X(z^{-1})$$

Since $\mathbf{x}$ is real valued, $X(z)$ is a polynomial in z with real coefficients and hence its zeros occur in conjugate pairs. Also, since $A(z) = A(z^{-1})$, if z_0 is a zero of $A(z)$, then z_0^{-1} is also a zero. Hence, the zeros of $A(z)$ appear in quadruples of the form $(z_0, z_0^\star, z_0^{-1}, z_0^{-\star})$. The extraction of $\mathbf{x}$ from $\mathbf{a}$, or equivalently $X(z)$ from $A(z)$, is known as spectral factorization (see [30]) and deals with the distribution of these quadruples between $X(z)$ and $X(z^{-1})$. For every quadruple $(z_0, z_0^\star, z_0^{-1}, z_0^{-\star})$, we can either assign $(z_0, z_0^\star)$ to $X(z)$ and $(z_0^{-1}, z_0^{-\star})$ to $X(z^{-1})$, or assign $(z_0^{-1}, z_0^{-\star})$ to $X(z)$ and $(z_0, z_0^\star)$ to $X(z^{-1})$, because of which, (3) can have upto $2^{n/2}$ non-equivalent solutions.

Hence, signal recovery from the autocorrelation (and hence Fourier transform magnitude) alone is not possible and additional information is necessary. In this work, we assume that the underlying signals

[§]In most applications, the measured autocorrelation corresponds to linear convolution and hence, in this work, we consider linear convolution when we refer to autocorrelation. We would like to note that (2) and (3) are equivalent when $\mathbf{x}$ is zero-padded with at least n zeros

[*]Throughout this work, when we refer to unique recovery, it is assumed to be up to an equivalent solution, i.e., up to time-shift, time-reversal and global sign.

are sparse, a property which is true in many applications of the phase retrieval problem. For example, astronomical imaging deals with the locations of the stars in the sky, electron microscopy deals with the density of atoms and so on. The sparse phase retrieval problem can be stated as

$$\begin{aligned} &\text{find} && \mathbf{x} \\ &\text{subject to} && |\mathbf{y}| = |\mathbf{F}\mathbf{x}| \\ & && ||\mathbf{x}||_0 << n \end{aligned} \tag{4}$$

or, equivalently,

$$\begin{aligned} &\text{find} && \mathbf{x} \\ &\text{subject to} && \mathbf{a} = \mathbf{x} \star \tilde{\mathbf{x}} \\ & && ||\mathbf{x}||_0 << n \end{aligned} \tag{5}$$

where $||.||_0$ is the l_0 norm.

1.1 Related Work

The phase retrieval problem has challenged researchers for decades, and considerable amount of research has been done. The Gerchberg-Saxton algorithm [8] was the first popular method to solve this problem when certain time domain constraints are imposed on the signal (sparsity can be considered as one such constraint). The algorithm starts by selecting a random Fourier phase, and then alternately enforces the time-domain constraints specific to the setup and the observed frequency-domain measurements. Fienup, in his seminal work [9], proposed a broad framework for such iterative algorithms. [10] provides a theoretical framework to understand these algorithms, which are in essence an alternating projection between a convex set and a non-convex set. The problem with such an approach is that convergence is often to a local minimum, hence chances of successful recovery are minimal.

Recently, attempts have been made by researchers to exploit the sparse nature of the underlying signals. [11] proposes an alternating projection-based heuristic to solve the sparse phase retrieval problem. [22] explores the traditional iterative algorithm with additional sparsity constraints. Semidefinite relaxation based heuristics were explored by several researchers (see [12], [13], [14]). In [16], a local-search method was explored to solve the sparsity constrained non-linear optimization problem. In [12, 21, 26, 23], the idea of using masks to obtain more information about the signal is explored.

We note that a considerable amount of literature is available on the "generalized" phase retrieval problem, which can be stated as

$$\begin{aligned} &\text{find} && \mathbf{x} \\ &\text{subject to} && |\mathbf{y}| = |\mathbf{A}\mathbf{x}| \end{aligned} \tag{6}$$

where $\mathbf{A}$ is a matrix with randomly chosen entries (see [15], [17], [18], [19], [20], [24], [25]). We would like to emphasize here that, while in appearance, (6) is similar to (2), (2) is more challenging due to the inherent structure of the measurement matrix. In particular, due to the trivial ambiguities time-shift and time-reversal, standard convex relaxation methods do not work. A detailed discussion is provided in Section 3.

2 Uniqueness

In this section, we show that most $O(n)$-sparse signals have distinct autocorrelations (and hence Fourier transform magnitudes).

Lemma 2.1. *If two distinct finite-length real-valued signals $\mathbf{x}_1$ and $\mathbf{x}_2$ have the same autocorrelation, then there exists finite-length real-valued signals $\mathbf{g}$ and $\mathbf{h}$ such that*

$$\mathbf{x}_1 = \mathbf{g} \star \mathbf{h} \quad \& \quad \mathbf{x}_2 = \mathbf{g} \star \tilde{\mathbf{h}}$$

where $\tilde{\mathbf{h}}$ is the time-reversed version of $\mathbf{h}$.

Proof. See Appendix. □

Definition: We say that a signal has *aperiodic* support if the locations of its non-zero components are not in an arithmetic progression (AP). For example:

(i) A signal which has $\{x_2, x_5, x_7\} \neq 0$ and the rest 0 has aperiodic support as $\{2, 5, 7\}$ are not in an AP.

(ii) A signal which has $\{x_2, x_5, x_8\} \neq 0$ and the rest 0 does not have aperiodic support as $\{2, 5, 8\}$ are in an AP.

(iii) A signal which has $\{x_1, x_2, x_3\} \neq 0$ and the rest 0 does not have aperiodic support as $\{1, 2, 3\}$ are in an AP.

Theorem 2.1. *Signals with aperiodic support can be uniquely recovered from their autocorrelation almost surely.*

Proof Outline. Consider the set of $(n-1)$-sparse signals with aperiodic support. The set of such signals is a manifold of dimension $n-1$. If two signals $\mathbf{x}_1$ and $\mathbf{x}_2$ from this set have the same autocorrelation, then from Lemma 2.1, we know that there exists two signals $\mathbf{g}$ and $\mathbf{h}$ such that

$$\mathbf{x}_1 = \mathbf{g} \star \mathbf{h} \;\; \& \;\; \mathbf{x}_2 = \mathbf{g} \star \tilde{\mathbf{h}}$$

Since one location in both $\mathbf{x}_1$ and $\mathbf{x}_2$ have a value zero, we have

$$\sum_r g_r h_{i-r} = 0 \;\; \& \;\; \sum_r g_r h_{j+r} = 0$$

for some i, j.

The set of such signals is a manifold of dimension $n-2$ (for details, see Appendix), which implies that almost all[†] signals in this set have a unique autocorrelation. The argument can be extended to the set of signals with aperiodic support (for details, see Appendix). □

Note that if a polynomial $X(z)$ is irreducible, it can be recovered uniquely from the polynomial $A(z) = X(z)X(z^{-1})$. This can be seen as a consequence of Lemma 2.1. In [27], it is shown that almost all signals in two or more dimensions can be uniquely recovered from their autocorrelation by using the fact that almost all polynomials in more than one variable are irreducible, i.e., the set of factorable polynomials in more than one variable is extremely small (measure zero). This result does not apply for one-dimensional signals, as all polynomials with real coefficients in one variable can be factored into linear and quadratic terms with real coefficients.

[†]The set of violations is measure zero

Theorem 2.1 in essence states that out of the $2^{n/2}$ possible solutions to (3), almost surely, only one of them has aperiodic support. This does not come as a surprise, as one can expect most solutions to have support $\{l_1, l_1+1....l_2-1, l_2\}$ for some $0 \leq l_1 \leq l_2 \leq n-1$, which is not aperiodic.

Note that sparse signals naturally meet the requirements of Theorem 2.1 unless the locations of the non-zero components are chosen periodically.[‡]

Theorem 2.2. *$O(n)$-sparse signals can be uniquely recovered from their Fourier transform magnitudes with probability $1-\delta$ for any $\delta > 0$ if*

(i) The support is chosen uniformly at random

(ii) $n > n_0(\delta)$

(iii) The signal values are chosen from a non-degenerate distribution

Proof. This is a direct consequence of Theorem 2.1[§]. □

Corollary 2.1. *Sparse signals can be uniquely recovered from their Fourier transform magnitudes with probability $1-\delta$ for any $\delta > 0$ if*

(i) Sparsity $k = O(n^{1/2-\epsilon})$

(ii) The support is chosen uniformly at random

(iii) $n > n_0(\delta, \epsilon)$

(iv) The signal values are chosen from a non-degenerate distribution

3 Two-stage Sparse Phase Retrieval (TSPR)

In this section, we first discuss the general approaches to solve such quadratically-constrained problems, talk about their drawbacks in the phase retrieval setup and then develop TSPR to solve (5).

A natural way to solve for sparse solutions in a set is by looking for points which minimize the l_0 norm. In particular, the sparse phase retrieval problem can be solved by

$$\begin{array}{ll} \text{minimize} & ||\mathbf{x}||_0 \\ \text{subject to} & \mathbf{a} = \mathbf{x} \star \tilde{\mathbf{x}} \end{array} \tag{7}$$

[‡]If the locations of the non-zero components are chosen periodically, the signal can be viewed as an oversampled copy of a non-sparse signal. The sparse phase retrieval problem reduces to the phase retrieval problem and hence cannot be uniquely solved.

[§]We would like to note that recent works [28, 29] provide guarantees assuming "collision free" support, a property which is true only for sparsities $O(n^{1/4-\epsilon})$ when the support is chosen at random.

However, it is well known that l_0 minimization is NP-hard. Also, the feasible set of (7) is clearly not convex. Suppose we embed the problem in a higher dimensional space by trying to solve for an $n \times n$ matrix $\mathbf{X}$ (a technique popularly known as *lifting*). (7) is equivalent to

$$\begin{aligned} &\text{minimize} && ||\mathbf{X}||_0 \\ &\text{subject to} && a_i = trace(\mathbf{A}_i\mathbf{X}) \qquad 0 \le i \le n-1 \\ & && \mathbf{X} \succcurlyeq 0 \quad \& \quad rank(\mathbf{X}) = 1 \end{aligned} \tag{8}$$

where the matrices $\mathbf{A}_i$ are given by

$$[\mathbf{A}_i]_{gh} = \begin{cases} 1 & \text{if} \quad |h-g| = i = 0 \\ 1/2 & \text{if} \quad |h-g| = i \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

Researchers have explored many convex relaxations to solve such problems. For example, l_1 minimization is a popular and powerful tool to promote sparse solutions [35]. [31] proposes the use of nuclear norm (or equivalently trace norm for positive semidefinite matrices) to promote low rank solutions. Since the solution we desire is both sparse and low rank, a natural approach would hence be to solve

$$\begin{aligned} &\text{minimize} && trace(\mathbf{X}) + \lambda||\mathbf{X}||_1 \\ &\text{subject to} && a_i = trace(\mathbf{A}_i\mathbf{X}) \qquad 0 \le i \le n-1 \\ & && \mathbf{X} \succcurlyeq 0 \end{aligned} \tag{9}$$

for some regularizer λ and hope that the resulting solution is both sparse and rank one. While this relaxation is a powerful tool when the measurement matrices are random (for example, the generalized phase retrieval setup in [36]), it fails in the phase retrieval setup.

This does not come as a surprise as the issue of trivial ambiguities due to time-shift and time-reversal is still unresolved. If $\mathbf{X}^\star = \mathbf{x}\mathbf{x}^T$ is the desired sparse solution, then $\tilde{\mathbf{X}} = \tilde{\mathbf{x}}\tilde{\mathbf{x}}^T$ where $\tilde{\mathbf{x}}$ is the flipped version of $\mathbf{x}$, $\mathbf{X}_i = \mathbf{x}_i\mathbf{x}_i^T$ where $\mathbf{x}_i$ is the signal obtained by time-shifting $\mathbf{x}$ by i units, and $\tilde{\mathbf{X}}_i = \tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^T$ where $\tilde{\mathbf{x}}_i$ is the signal obtained by time-shifting $\tilde{\mathbf{x}}$ by i units are also feasible points with the same objective function value as $\mathbf{X}^\star$. Since (9) is a convex problem, any convex combination of these equivalent solutions are also feasible, and have an objective function value less than or equal to that of $\mathbf{X}^\star$. Hence the optimizer is neither sparse nor rank one. One approach to break this symmetry would be to solve a weighted l_1 minimization problem (for example, see [32]), which introduces a bias towards a particular equivalent solution. Unfortunately, this approach does not help in the phase retrieval setup.

[12] proposes a heuristic by using log-det function as a surrogate for the rank function [34] and solves

$$\begin{aligned} &\text{minimize} && log(det(\mathbf{X})) \\ &\text{subject to} && a_i = trace(\mathbf{A}_i\mathbf{X}) \qquad 0 \le i \le n-1 \\ & && \mathbf{X} \succcurlyeq 0 \end{aligned} \tag{10}$$

by recasting the problem as minimization of a concave function over a convex domain and using gradient descent to solve for a local minimum. Many other iterative heuristics were proposed to solve (9). For example, [13] iteratively reduces the solution space by calculating bounds on the support of the signal, [37] explores reweighted minimization (see [33]) where the weights are chosen based on the solution of the previous iteration. While these methods enjoy empirical success, no guarantees were provided for their behavior.

3.1 Recovery with known support

The time-shift and time-reversal ambiguities stem from the fact that the support of the signal is not known. Therefore, let us momentarily assume that we somehow know the support of the signal (denoted from now on by V, which is the set of locations of the non-zero components of $\mathbf{x}$), (9) can be reformulated as

$$\begin{array}{ll} \text{minimize} & trace(\mathbf{X}) + \lambda ||\mathbf{X}||_1 \\ \text{subject to} & a_i = trace(\mathbf{A}_i \mathbf{X}) \qquad 0 \le i \le n-1 \\ & X_{ij} = 0 \quad \text{if} \quad \{i,j\} \notin V \\ & \mathbf{X} \succcurlyeq 0 \end{array} \tag{11}$$

Figure 1 plots the probability of the solution to (11) being rank one against various sparsities k for $n = 64, 128$. For a given n and k, the k non-zero locations were chosen uniformly at random and the signal values in the support were chosen from an i.i.d Gaussian distribution. It can be observed that (11) recovers the underlying signal with very high probability as long as the sparsity satisfies $k \lesssim n/2$[¶]. This observation suggests a two-stage algorithm: one where we recover the support of the signal first and then use it to solve (11).

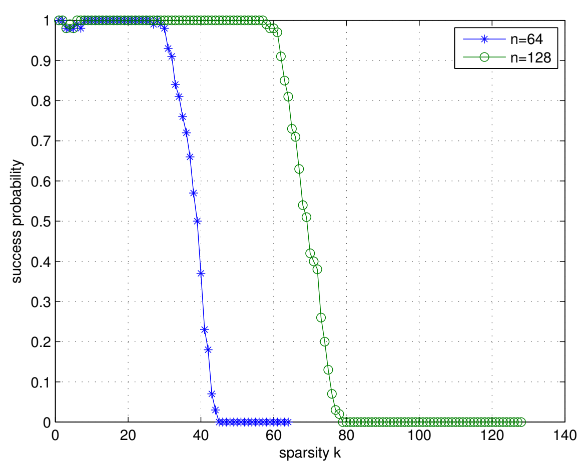


Figure 1: Probability of successful signal recovery of (11)

[¶]This is an empirical observation. In this work, we provide proof only up to $k = O(n^{1/2-\epsilon})$ regime.

In the remaining part of this paper, we develop TSPR, which provably recovers $O(n^{1/2-\epsilon})$-sparse signals from their Fourier transform magnitudes with arbitrarily high probability.

Algorithm 1 Two-stage Sparse Phase Retrieval (TSPR)

Input: Autocorrelation (i.e., Fourier transform magnitude)
Output: Sparse signal which satisfies the measurements

(i) Recover the support of the signal using Algorithm 2 or 3

(ii) Recover the signal values in the support using Algorithm 4 or 5

Theorem 3.1. *TSPR can recover $O(n^{1/2-\epsilon})$-sparse signals from their Fourier transform magnitudes in quadratic-time with probability $1-\delta$ for any $\delta > 0$ if*

(i) V chosen uniformly at random from $\{0, 1, ..., n-1\}$

(ii) $n > n_0(\epsilon, \delta)$

(iii) The signal values are chosen from a non-degenerate distribution

Proof. This is a direct consequence of Theorem 4.1 and 5.1. □

4 Support Recovery

In this section, we show that the support of sparse signals can be uniquely recovered from the support of their autocorrelation with arbitrarily high probability if their sparsity is $O(n^{1/2-\epsilon})$ and develop two recovery algorithms. The first algorithm (Algorithm 2) uses a combinatorial approach, and the second (Algorithm 3) is convex optimization-based. We also provide the necessary recovery guarantees for the combinatorial approach. Also, we relate this recovery problem to a problem well known in literature as the turnpike problem [40, 45, 48], which is the problem of recovering integer sets from their pairwise distance sets. The proposed algorithms can hence be used to solve the "sparse" turnpike problem.

It is often useful to be able to reconstruct the support of the signal V from the (support of the) autocorrelation W [38]. In many applications (e.g, astronomy), the signal's support is the desired information. In other applications, support knowledge makes the signal reconstruction process using available techniques significantly easier [39].

We will assume that if $a_i = 0$, then no two elements in $\mathbf{x}$ are separated by a distance i, i.e.,

$$a_i = 0 \Rightarrow x_j x_{i+j} = 0 \ \forall \ j$$

This is a very weak assumption and holds with probability one if the non-zero components of the signal are chosen from a non-degenerate distribution. With this assumption, the support recovery problem can be stated as

$$\begin{aligned} &\text{find} && V \\ &\text{subject to} && \{|i-j| \mid (i,j) \in V\} = W \end{aligned} \tag{12}$$

Note that V is a set of integers, and W is its pairwise distance set.

4.1 Turnpike Problem

Turnpike problem† is the problem of reconstructing a set of integers from the set of their pairwise distances. For example, consider the set $V = \{2, 5, 13, 31, 44\}$. Its pairwise distance set is given by $W = \{0, 3, 8, 11, 13, 18, 26, 29, 31, 39, 42\}$. Turnpike problem deals with the recovery of the set V from the set W‡.

This problem has received considerable attention in computational biology. Over the last few years, there has been a lot of interest in DNA restriction site analysis. A DNA strand is a string on the letters $\{A, T, G, C\}$. Unfortunately, the DNA string cannot be explicitly observed and in order to map it, biochemical techniques which provide indirect information have been developed.

When a particular restriction enzyme is added to a DNA solution, the DNA is cut at particular restriction sites. For example, the enzyme *EcoRI* cuts at locations of the pattern GAATTC. The goal of restriction site analysis is to determine the locations of every site for a given enzyme. In order to do this, a batch of DNA is exposed to a restriction enzyme in limited quantity so that fragments of all possible lengths exist (see Figure 2). Using gel electrophoresis, the fragment lengths can be measured.

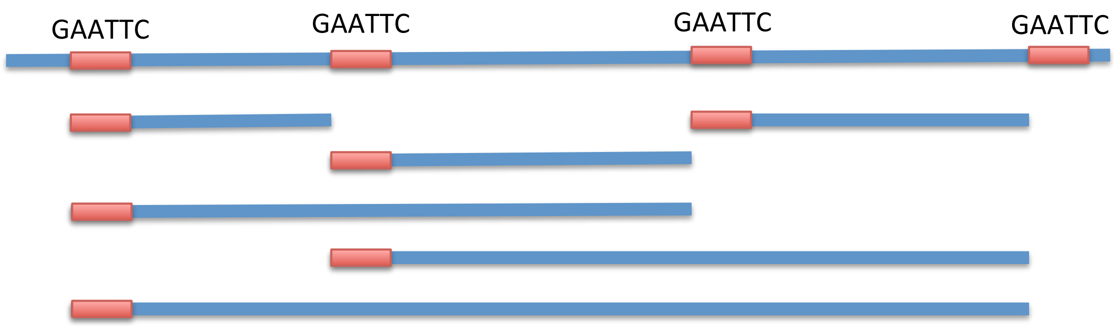


Figure 2: Partial Digest Problem

Recovering the locations of the restriction sites from the measured fragment lengths is known as the partial digest problem (a.k.a turnpike problem, see [45]). The locations of the restriction sites correspond to the set of integers V, and the measured fragment lengths correspond to the set of pairwise distances W.

[42] derives some properties of sets which have the same pairwise distance set, similar to Lemma 2.1 in flavor. [43] develops a polynomial factorization-based algorithm with complexity $O(n^d)$, where d is the

---

†Many papers consider the problem of recovering a set of integers from the multiset of their pairwise distances, i.e., multiplicity of every pairwise distance is also known. In this work, we provide a solution without using multiplicity information, which can be trivially extended to the case where multiplicity information is known by discarding the extra information

‡If V has a pairwise distance set W, then sets $c \pm V$ also have the same pairwise distance set W for any integer c. These solutions are considered equivalent, and in all the applications it is considered good enough if any equivalent solution, i.e., up to linear translation and flipping, is recovered.

largest pairwise distance. [40] proposes a backtracking-based algorithm to solve the turnpike problem. The algorithm starts with an empty set and adds elements to the solution by using the idea that the largest "unexplained" pairwise distance at every step has to include at least one end and proceeds by guessing the end, backtracking if there are contradictions. The algorithm is known to have a worst case exponential $O(2^n nlogn)$ complexity [44]. [46] proposes a dynamic programming-based algorithm, which is worst case exponential complexity too. [47] develops a polynomial time algorithm by using special measurement systems to obtain more information about the signal. [45] provides a very good summary of the existing algorithms.

In this work, we consider the turnpike problem with the additional assumption that the underlying integer set is "sparse", i.e., sufficiently spaced, and provide a $O(k^4)$ complexity algorithm which can recover the integer set with arbitrarily high probability. Specifically, suppose $V = \{v_0, v_1, ..., v_{k-1}\}$ is a set of k integers and $W = \{w_0, w_1, ..., w_{K-1}\}$ is its pairwise distance set[‡].

Theorem 4.1. *(Support of the signal) V can be recovered uniquely (upto linear shift and reversal) from (support of the autocorrelation) W using a $O(k^4)$-complexity (equivalently $O(n^2)$-complexity) algorithm (Algorithm 2) with probability greater than $1-\delta$ for any $\delta > 0$ if*

(i) V *chosen uniformly at random from* $\{0, 1, ..., n-1\}$

(ii) $k = O(n^{1/2-\epsilon})$

(iii) $n > n_0(\epsilon, \delta)$

Proof. In the next part of this paper, we develop a combinatorial algorithm (Algorithm 2). The proof of this theorem is constructive, i.e., we prove the correctness of the various steps involved in Algorithm 2. Refer to Appendix for details. □

4.2 Combinatorial Algorithm

In order to overcome the trivial ambiguity of linear shift and reversal, we attempt to recover the equivalent solution set $U = \{u_0, u_1, ..., u_{k-1}\}$ defined as follows

$$U = \begin{cases} V - v_0 & \text{if} \quad v_1 - v_0 < v_{k-1} - v_{k-2} \\ v_{k-1} - V & \text{otherwise} \end{cases}$$

i.e., the equivalent solution set U we attempt to recover has the following properties:

(i) $u_0 = 0$

(ii) $u_1 - u_0 < u_{k-1} - u_{k-2}$

Let $u_{ij} = u_j - u_i$ for $0 \leq i \leq j \leq k-1$. With this definition, $W = \{u_{ij} : 0 \leq i \leq j \leq k-1\}$ and $U = \{u_{0j} : 0 \leq j \leq k-1\}$. Note that $U \subseteq W$.

[‡]The elements of V and W are assumed to be in ascending order without loss of generality for convenience of notation, i.e., $v_0 < v_1 < .... < v_{k-1}$ and $w_0 < w_1 < .... < w_{K-1}$

Algorithm 2 Support Recovery: Combinatorial Algorithm

Input: Pairwise distance set W

Output: Integer set U which realizes W

1. Infer u_{01} from W
2. Construct the set $W_1 = W + u_{01}$
3. Construct the graph $G((\{0\} \cup (W \cap W_1)), W)$ and infer $\{u_{0p} : 1 \leq p \leq t = log(k)\}$ by Graph step.
4. Construct the set $W_p = W + u_{0p}$ for $1 \leq p \leq t$
5. Calculate $U_t = W \cap \left(\bigcap_{p=1}^{t} W_p\right)$
6. Recover $U = \{u_{0p} : 0 \leq p \leq t-1\} \cup U_t$

4.2.1 Intersection Step

The key idea of this step can be summarized as follows: suppose we know the value of u_{0p} for some p, if U_p and W_p are defined as

$$U_p = \{u_{0j} : p \leq j \leq k-1\} \quad \& \quad W_p = W + u_{0p}$$

then we have

$$U_p \subseteq W \cap W_p$$

The idea can be extended to multiple intersections. Suppose we know $\{u_{0p} : 1 \leq p \leq t\}$, we can construct $\{W_p : 1 \leq p \leq t\}$ and see that

$$U_t \subseteq W \cap \left(\bigcap_{p=1}^{t} W_p\right)$$

4.2.2 Graph Step

For an integer set U whose pairwise distance set is W, consider the set $Z = \{z_0, z_1, \dots z_{|Z|-1}\}$ such that $U \subseteq Z \subseteq W$. Construct a graph $G(Z, W)$ with $|Z|$ vertices (each vertex corresponding to an element of Z) such that there exists an edge between z_i and z_j iff the following two conditions are satisfied

(i) $\forall z_g, z_h \in Z, z_g - z_h \neq z_i - z_j \quad \text{unless} \quad (i,j) = (g,h)$

(ii) $z_i - z_j \in W$

i.e., there exists an edge between two vertices if their corresponding pairwise distance is unique and belongs to W. For example, consider the integer set $U = \{0, 10, 15, 50, 80\}$ whose pairwise distance set is $W = \{0, 5, 10, 15, 30, 35, 40, 50, 65, 70, 80\}$. Consider the set $Z = \{0, 10, 15, 35, 40, 50\}$. The graph

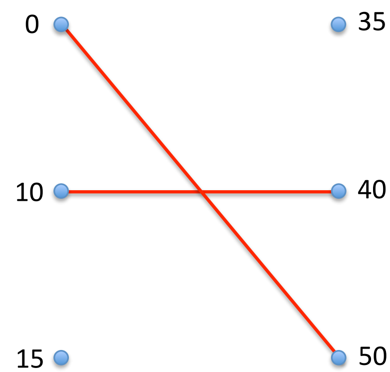


Figure 3: $G(Z, W)$ for $Z = \{0, 10, 15, 35, 40, 50\}$ and $W = \{0, 5, 10, 15, 30, 35, 40, 50, 65, 70, 80\}$

$G(Z, W)$ looks as shown in Figure 3. Note that there exists an edge between 10 and 40 as it is the only pair of integers with difference 30 and $30 \in W$, there doesn't exist an edge between 0 and 40 as there is another pair $\{10, 50\}$ with difference 40 and so on.

The main idea of this step is as follows: suppose we draw a graph $G(Z, W)$ where $U \subseteq Z \subseteq W$. If there exists an edge between a pair of integers $\{z_i, z_j\} \in Z$, then $\{z_i, z_j\} \in U$. This holds because if $\{z_i, z_j\} \notin U$, then since $z_i - z_j \in W$ there has to be another pair of integers in U (and hence in Z) which have a pairwise distance $z_i - z_j$. This would contradict the fact that an edge exists between z_i and z_j in $G(Z, W)$.

Remark: For better performance, Algorithm 2 can be modified slightly (see Algorithm 7 in Appendix).

4.3 Convex Algorithm

Since the support recovery problem (12) has quadratic constraints, we can relax them into a set of convex constraints by using semidefinite relaxation which has shown great promise in some non-convex quadratically constrained problems. Suppose we define $\mathbf{s} = (s_0, s_1, \ldots. s_{n-1})$ as the indicator function of the signal, i.e.,

$$s_i = \begin{cases} 1 & \text{if } x_i \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

The support recovery problem can be stated as

$$\begin{aligned} &\text{find} && \mathbf{s} && && (13) \\ &\text{subject to} && \sum_i s_i s_{i+j} > 0 && \text{iff } a_j \neq 0 \\ & && s_i \in \{0, 1\} && 0 \leq i \leq n-1 \end{aligned}$$

By lifting the problem, we can rewrite (13) as

$$\begin{aligned}
&\text{find} && \mathbf{S} \\
&\text{subject to} && \sum_i S_{i,i+j} > 0 \qquad \text{iff } a_j \neq 0 \\
& && S_{ij} \in \{0,1\} \qquad 0 \leq i,j \leq n-1 \\
& && \mathbf{S} \succcurlyeq 0 \quad \& \quad rank(\mathbf{S}) = 1
\end{aligned} \tag{14}$$

In order to obtain the tightest convex relaxation, $\mathbf{S}$ is allowed to be positive semidefinite as it is the smallest convex set containing all rank one matrices. Its entries are allowed to be in the interval $[0,1]$, which is the tightest convex relaxation for binary variables. Suppose we also know the sparsity k of the signal, the trace of $\mathbf{S}$ is given by

$$trace(\mathbf{S}) = \sum_i s_i^2 = \sum_i s_i = k$$

Also, for row i (or column i),

$$\sum_i S_{ij} \left(\text{or } \sum_i S_{ji} \right) = \sum_i s_i s_j = s_j \sum_i s_i = k s_j = k s_j^2 = k S_{jj}$$

Note that we still haven't resolved the issue of trivial ambiguity. Since equivalent solutions also satisfy all the constraints, a random matrix $\mathbf{\Theta}$ is used as a weight matrix to bias the solution towards one of the equivalent solutions. The support recovery problem hence can be relaxed to (15).

Algorithm 3 Support Recovery: Convex Algorithm

Input: Autocorrelation $\mathbf{a}$ and sparsity k of the signal
Output: Indicator function $\mathbf{s}$ of the signal

(i) Obtain $\mathbf{S}^\star$ by solving

$$\begin{aligned}
&\text{minimize} && trace(\mathbf{\Theta S}) \\
&\text{subject to} && trace(\mathbf{S}) = k \\
& && \sum_i S_{ij} = k S_{jj} \quad \& \quad \sum_j S_{ij} = k S_{ii} \qquad 0 \leq i,j \leq n-1 \\
& && \sum_i S_{i,i+j} > 0 \text{ iff } a_j \neq 0 \\
& && 0 \leq S_{ij} \leq 1 \qquad 0 \leq i,j \leq n-1 \\
& && \mathbf{S} \succcurlyeq 0
\end{aligned} \tag{15}$$

(ii) Return $\mathbf{s}$, where $\mathbf{S}^\star = \mathbf{s}\mathbf{s}^T$

Note that Algorithm 3 requires the knowledge of k, which is not available in general. Approximate bounds can be easily calculated by looking at the number of non-zero components in the autocorrelation if $k = O(n^{1/2-\epsilon})$. These bounds can instead be used as a viable alternative by relaxing the corresponding constraints to account for approximation.

5 Signal Value Recovery

In this section, we develop two algorithms for recovering the signal values once the support is known. The first algorithm (Algorithm 4) uses a combinatorial approach, and the second (Algorithm 5) is convex optimization-based. We also provide the necessary recovery guarantees for the proposed algorithms.

The signal value recovery problem once the support is known can be stated as

$$\begin{aligned} &\text{find} && \mathbf{x} \\ &\text{subject to} && \mathbf{a} = \mathbf{x} \star \tilde{\mathbf{x}} \\ &&& x_i = 0 \quad \text{if} \quad i \notin U \qquad 0 \leq i \leq n-1 \end{aligned} \tag{16}$$

Theorem 5.1. *Signal values can be uniquely extracted from the autocorrelation once the signal's support is known using a $O(n)$ or $O(k^6)$-complexity algorithm (Algorithm 4 or 5 respectively) with probability $1-\delta$ for any $\delta > 0$ if*

(i) V chosen uniformly at random from $\{0, 1, ..., n-1\}$

(ii) $k = O(n^{1/2-\epsilon})$

(iii) $n > n_0(\epsilon, \delta)$

Proof. The proof is constructive (Algorithm 4 or 5), and is a direct consequence of Lemmas 5.1 and 5.2 or Lemma 5.3. For algorithm 4, note that triangle finding can be done in constant-time due to very high edge density of $H(V)$. □

5.1 Combinatorial Algorithm

Suppose $V = (v_0, v_1, ...v_{k-1})$ is the support of the signal. Construct a graph $H(V)$ with k vertices (each vertex corresponding to an element of V) such that there exists an edge between every pair of vertices v_i and v_j which satisfy the following condition

$$\forall v_g, v_h \in V, \quad v_g - v_h \neq v_i - v_j \text{ unless} \quad (i,j) = (g,h)$$

The key idea is that if there exists an edge between v_i and v_j, then

$$x_{v_i} x_{v_j} = a_{|v_i - v_j|}$$

i.e., if an edge exists between two vertices, then the product of the corresponding signal values is known. This is true because no other term can contribute to $a_{|v_i - v_j|}$ in the autocorrelation.

Lemma 5.1. *Suppose the graph $H(V)$ is connected and has a triangle, then the signal can be extracted uniquely from the autocorrelation up to a global sign.*

Proof. Without loss of generality, let the induced subgraph of $\{v_0, v_1, v_2\}$ be a triangle. In other words, we know the pairwise products $(x_{v_1}x_{v_2})$, $(x_{v_2}x_{v_0})$ and $(x_{v_1}x_{v_0})$. Note that

$$\frac{(x_{v_0}x_{v_2})(x_{v_1}x_{v_0})}{(x_{v_1}x_{v_2})} = {x_{v_0}}^2, \quad x_{v_1} = \frac{(x_{v_1}x_{v_0})}{x_{v_0}} \quad \& \quad x_{v_2} = \frac{(x_{v_2}x_{v_0})}{x_{v_0}}$$

from which $(x_{v_0}, x_{v_1}, x_{v_2})$ can be recovered up to a global sign. Note that if there is an edge between v_i and v_j, and if one of x_{v_i} or x_{v_j} is known, the other can be recovered. Since the graph $H(V)$ is connected, with the knowledge of x_{v_0} up to a sign, all the signal values can be recovered up to a global sign. □

Algorithm 4 Signal Value Recovery: Combinatorial Algorithm

Input: Autocorrelation $\mathbf{a}$ and signal's support V (or equivalently U)
Output: Signal $\mathbf{x}$ up to global sign with the desired support and autocorrelation

(i) Construct the graph $H(V)$

(ii) Find a triangle in $H(V)$ with vertices (say) $v_{i_1}, v_{i_2}, v_{i_3}$, and calculate $x_{v_{i_1}}, x_{v_{i_2}}, x_{v_{i_3}}$ up to a global sign.

(iii) Find a spanning tree in $H(V)$ and calculate $\mathbf{x}$ up to a global sign.

Lemma 5.2. *The graph $H(V)$ is connected and has a triangle with probability $1-\delta$ for any $\delta > 0$ if*

(i) V chosen uniformly at random from $\{0, 1, ..., n-1\}$

(ii) $k = O(n^{1/2-\epsilon})$

(iii) $n > n(\epsilon, \delta)$

Proof. See Appendix. □

5.2 Convex Algorithm

Note that (16) is a non-convex problem as the autocorrelation constraints are non-convex. We can relax (16) to obtain a convex problem (Algorithm 5) using semidefinite relaxation, as we saw earlier in (11).

Lemma 5.3. *The program (17) returns a rank one solution with probability $1-\delta$ for any $\delta > 0$ if*

(i) V chosen uniformly at random from $\{0, 1, ..., n-1\}$

(ii) $k = O(n^{1/2-\epsilon})$

(iii) $n > n_0(\epsilon, \delta)$

Proof. See Appendix. □

Algorithm 5 Signal Value Recovery: Convex Algorithm

Input: Autocorrelation $\mathbf{a}$ and signal's support V (or equivalently U)
Output: Signal $\mathbf{x}$ upto global sign with the desired support and autocorrelation

(i) Obtain $\mathbf{X}^\star$ by solving

$$\begin{aligned} \text{minimize} \quad & \mathit{trace}(\mathbf{X}) \\ \text{subject to} \quad & a_i = \mathit{trace}(\mathbf{A}_i\mathbf{X}) \qquad 0 \le i \le n-1 \\ & X_{ij} = 0 \quad \text{if} \quad \{i,j\} \notin V \\ & \mathbf{X} \succcurlyeq 0 \end{aligned} \tag{17}$$

(ii) Return $\mathbf{x}$, where $\mathbf{X}^\star = \mathbf{x}\mathbf{x}^T$

6 Stability

In this section, we consider the impact of measurement noise on TSPR (the measured autocorrelation is corrupted with additive noise $\mathbf{n}$) and provide stability guarantees.

$$\mathbf{a} = (\mathbf{x} \star \tilde{\mathbf{x}}) + \mathbf{n}$$

Algorithm 6 Two-stage Sparse Phase Retrieval: Noisy Setup

Input: Autocorrelation $\mathbf{a}$ of the signal $\mathbf{x}$, η such that $||\mathbf{n}||_\infty < \eta$
Output: Signal $\hat{\mathbf{x}}$ upto global sign such that $||\mathbf{x} - \hat{\mathbf{x}}||_2^2 \le 4k\eta$

(i) Threshold the autocorrelation values using η to obtain the support of the autocorrelation

(ii) Solve for the support of the signal using Algorithm 2 or 3

(iii) Obtain $\mathbf{X}^\star$ by solving

$$\begin{aligned} \text{minimize} \quad & \mathit{trace}(\mathbf{X}) \\ \text{subject to} \quad & |a_i - \mathit{trace}(\mathbf{A}_i\mathbf{X})| < \eta \qquad 0 \le i \le n-1 \\ & X_{ij} = 0 \quad \text{if} \quad \{i,j\} \notin U \\ & \mathbf{X} \succcurlyeq 0 \end{aligned} \tag{18}$$

(iv) Return $\hat{\mathbf{x}}$, where $\mathbf{X}^\star = \hat{\mathbf{x}}\hat{\mathbf{x}}^T$

Theorem 6.1. *Suppose $\mathbf{n}$ satisfies $||\mathbf{n}||_\infty < \eta$. If $\hat{\mathbf{x}}$ is the solution of Algorithm 6, then*

$$||\mathbf{x} - \hat{\mathbf{x}}||_2^2 \le 4k\eta$$

with probability $1-\delta$ for any $\delta > 0$ if *(i) Support chosen uniformly at random from* $\{0, 1, ..., n-1\}$

(ii) $k = O(n^{1/2-\epsilon})$

(iii) $n > n_0(\epsilon, \delta)$

(iv) $\eta < |a_{w_i}|/2 \qquad 0 \leq i \leq K-1$

Proof. See Appendix □

7 Numerical Simulations

In this section, we demonstrate the performance of TSPR using numerical simulations. Various values of signal lengths n and sparsities k are used in the experiments. The procedure is as follows: for a given n and k, the k locations of the non-zero components were chosen uniformly at random. The signal values in the chosen support were drawn from an i.i.d standard normal distribution.

7.1 Support Recovery

In the first set of experiments, we evaluate the performance of the support recovery step of TSPR. In order to understand the various steps of the combinatorial algorithm (Algorithm 2), we provide the working details for a particular example: let $V = \{2, 5, 7, 19, 56, 108, 113, 116, 146, 184\}$. Its pairwise distance set is given by

$$W = \{0, 2, 3, 5, 8, 12, 14, 17, 30, 33, 37, 38, 49, 51, 52, 54, 57, 60, 68, 71, 76, 89, 90, 94, 97, 101,$$
$$103, 106, 108, 109, 111, 114, 127, 128, 139, 141, 144, 165, 177, 179, 182\}$$

We can infer $u_{01} = 182 - 179 = 3$ from W. Construct $W_1 = W + u_{01}$ and calculate $W \cap W_1$.

$$W \cap W_1 = \{3, 5, 8, 17, 33, 52, 54, 57, 60, 71, 97, 106, 109, 111, 114, 144, 182\}$$

Construct $G((\{0\} \cup (W \cap W_1)), W)$ to see that

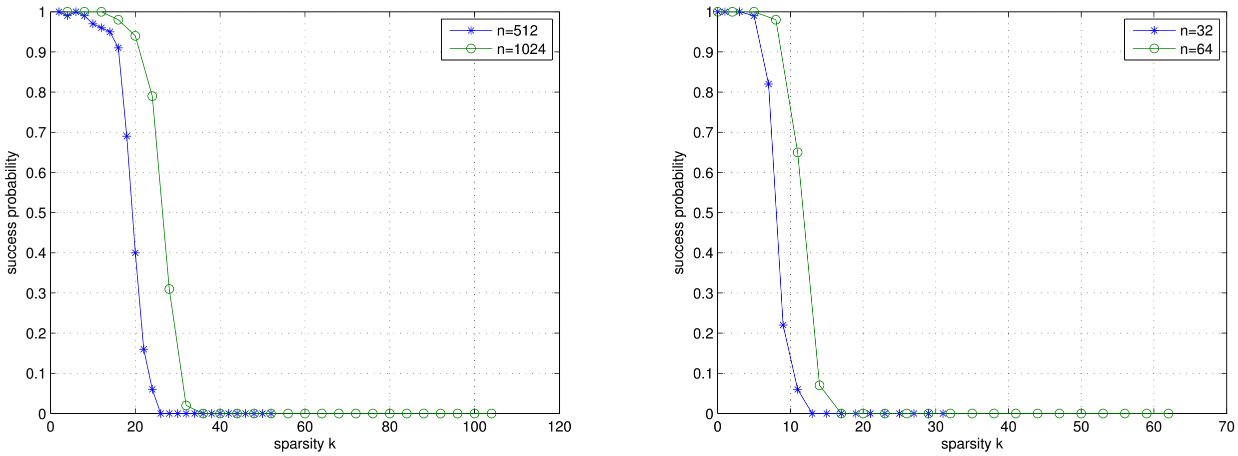
$$\{182\} \leftrightarrow \{5\} \quad \& \quad \{182\} \leftrightarrow \{17\}$$

from which we can infer $u_{02} = 5$ and $u_{03} = 17$. Construct $W_2 = W + u_{02}$, $W_3 = W + u_{03}$ and calculate $\left(\bigcap_{p=0}^{3} W_p\right)$

$$\left(\bigcap_{p=0}^{3} W_p\right) = \{54, 106, 111, 114, 144, 182\}$$

Calculate $U = \{u_{0p} : 0 \leq p \leq 3\} \bigcup \left(\bigcap_{p=0}^{3} W_p\right)$

$$U = \{0, 3, 5, 17, 54, 106, 111, 114, 144, 182\}$$



(a) Algorithm 2 (b) Algorithm 3

Figure 4: Probability of successful support recovery

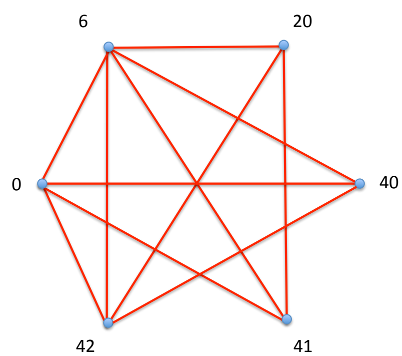


Figure 5: The graph $H(V)$ for $V = \{0, 1, 7, 9, 10, 27, 43, 47\}$

which is the desired solution.

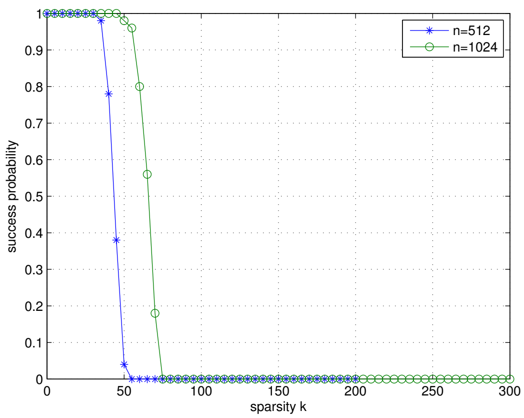
Figure 4(a) plots the probability of successful support recovery of Algorithm 2 for $n = 512, 1024$. The observations are consistent with Theorem 4.1, and also suggest that the $O(n^{1/2-\epsilon})$ bound is tight.

Figure 4(b) plots the probability of successful support recovery of Algorithm 3 for $n = 32, 64$. Simulations strongly suggest a similar $k = O(n^{1/2-\epsilon})$ bound.

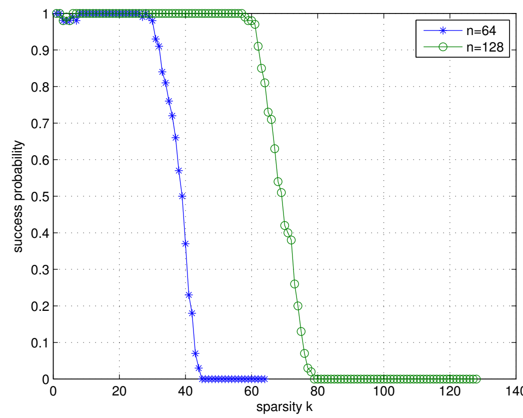
7.2 Signal Recovery

In the second set of experiments, we evaluate the performance of the signal recovery step of TSPR.

Figure 5 shows the graph $H(V)$ for a randomly generated support set with $k = 6$ and $n = 64$. Observe that $H(V)$ is dense, in accordance with Lemma 9.13. This density can be utilized to make the combinatorial algorithm robust in presence of noise, by calculating signal values using many triangles to prevent error propagation.



(a) Algorithm 4



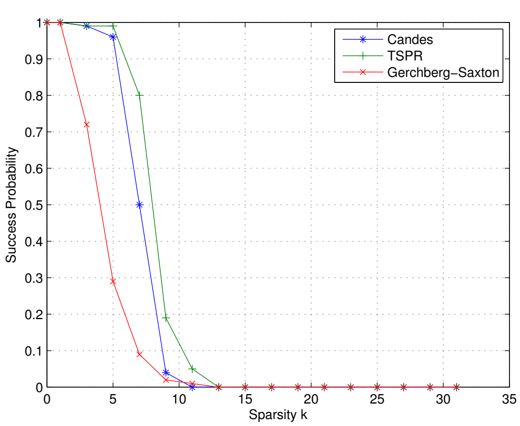
(b) Algorithm 5

Figure 6: Probability of successful signal recovery

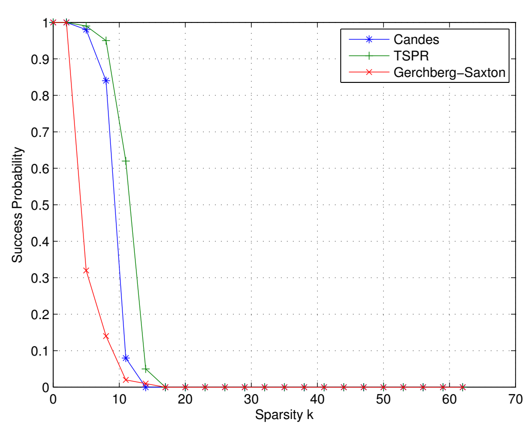
The probability of successful recovery of Algorithm 4 is shown in Figure 6(a), the $k = O(n^{1/2-\epsilon})$ behavior can be clearly observed. The performance of Algorithm 5 was already demonstrated in Figure 1, and is provided here for completeness. The recovery is successful with very high probability for sparsities k up to $\sim n/2$. In this work, the recovery guarantee for this algorithm is provided only up to $k = O(n^{1/2-\epsilon})$.

7.3 Performance of TSPR

Figures 7(a), 7(b) combine the results mentioned above and plot the performance of TSPR for $n = 32, 64$ and various sparsities. A comparison is provided with other popular algorithms (Gerchberg-Saxton [8], Candes [12]).



(a) n=32



(b) n=64

Figure 7: Performance of Phase Retrieval Algorithms

8 Summary

In this work, we considered the sparse phase retrieval problem. We showed that most $O(n)$-sparse signals have distinct Fourier transform magnitudes (Theorem 2.2). We then demonstrated a two-stage algorithm TSPR which provably (Theorem 3.1) recovered $O(n^{1/2-\epsilon})$-sparse signals from their Fourier transform magnitudes in quadratic-time. In the first stage, we recovered the support of the signal and in the second stage, we recovered the signal values using the support knowledge. For both stages, we provided a combinatorial and a convex-relaxation algorithm.

An important problem that remains open is whether one can improve the $O(n^{1/2-\epsilon})$ bound using fast algorithms.

9 Appendix

9.1 Uniqueness Proofs

9.1.1 Proof of Lemma 2.1

Let $X_1(z)$, $X_2(z)$, $G(z)$ and $H(z)$ be the z-transforms of the signals $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{g}$ and $\mathbf{h}$ respectively. Since $\mathbf{x}_1$ and $\mathbf{x}_2$ have the same autocorrelation, we have

$$A(z) = X_1(z)X_1(z^{-1}) = X_2(z)X_2(z^{-1})$$

where $A(z)$ is the z-transform of the autocorrelation of $\mathbf{x}_1$ and $\mathbf{x}_2$. For every quadruple $(z_0, z_0^\star, z_0^{-1}, z_0^{-\star})$ which are zeros of $A(z)$, $(z_0, z_0^\star)$ has to be assigned to $X_1(z)$ or $X_1(z^{-1})$, and $X_2(z)$ or $X_2(z^{-1})$. Let $P_1(z)$, $P_2(z)$ and $P_3(z)$ be the polynomials constructed from such conjugate pairs of zeros which are assigned to $(X_1(z), X_2(z))$ and $(X_1(z), X_2(z^{-1}))$ and $(X_1(z^{-1}), X_2(z))$ respectively. Note that $P_2(z) = P_3(z^{-1})$. We have

$$X_1(z) = P_1(z)P_2(z)$$
$$X_2(z) = P_1(z)P_3(z) = P_1(z)P_2(z^{-1})$$

and hence $X_1(z)$ and $X_2(z)$ can be written as

$$X_1(z) = G(z)H(z) \qquad X_2(z) = G(z)H(z^{-1})$$

where $G(z) = P_1(z)$ and $H(z) = P_2(z)$, or equivalently

$$\mathbf{x}_1 = \mathbf{g} \star \mathbf{h} \qquad \mathbf{x}_2 = \mathbf{g} \star \tilde{\mathbf{h}}$$

in the time domain as the z-transform of $\tilde{\mathbf{h}}$ is $H(z^{-1})$.

9.2 Proof of Theorem 2.1

Lemma 9.1. *If $f : \mathcal{A} \to \mathcal{B}$ is a map from $\mathcal{A}$ to $\mathcal{B}$, where $\mathcal{A}$ is a manifold of dimension d_a and $\mathcal{B}$ is a manifold of dimension d_b, then the image of f is measure zero in $\mathcal{B}$ if $d_a < d_b$.*

Proof. See [27]. $\square$

Lemma 9.2. *Let $\mathcal{S}_1$ be the set of k-sparse signals with aperiodic support which can be written as $\mathbf{g} * \mathbf{h}$ where both $\mathbf{g}$ and $\mathbf{h}$ have periodic support and $\mathcal{S}_2$ be the set of k-sparse signals with aperiodic support which can be written as $\mathbf{g} * \mathbf{h}$ where either $\mathbf{g}$ or $\mathbf{h}$ have aperiodic support. $\mathcal{S}_2$ has a dimension strictly lesser than $\mathcal{S}_1$ for all k.*

Proof. Note that the set of k-sparse signals is a manifold of dimension k. The sparsity constraints can be viewed as a set of bilinear equations in $\mathbf{g}$ and $\mathbf{h}$ [21]. Both $\mathcal{S}_1$ and $\mathcal{S}_2$ satisfy these constraints and the dimension reduction due to sparsity is identical. However, $\mathcal{S}_2$ also requires at least one of the components in either $\mathbf{g}$ or $\mathbf{h}$ to be zero, because of which there is a further reduction in dimension by atleast 1. $\square$

In essence, Lemma 9.2 suggests that while considering the set of k-sparse signals with aperiodic support, if we consider only those k-sparse signals which can be formed by convolving two signals $\mathbf{g}$ and $\mathbf{h}$ which have periodic support, we are losing only a measure zero set of signals.

Lemma 9.3. *Suppose $\mathbf{g}$ and $\mathbf{h}$ have periodic support. Let $\mathcal{T}_1$ be the set of $\mathbf{g}$ and $\mathbf{h}$ such that $\mathbf{g} * \mathbf{h}$ has aperiodic support with sparsity k and let $\mathcal{T}_2$ be the set of $\mathbf{g}$ and $\mathbf{h}$ such that $\mathbf{g} * \mathbf{h}$ has aperiodic support with sparsity k and $\mathbf{g} * \tilde{\mathbf{h}}$ has aperiodic support. $\mathcal{T}_2$ has a dimension strictly less than $\mathcal{T}_1$.*

Proof. The sparsity constraints can be viewed as a set of bilinear equations in $\mathbf{g}$ and $\mathbf{h}$ [21] and result in identical dimension reduction for both $\mathcal{T}_1$ and $\mathcal{T}_2$. If the sets $\mathbf{g}$ and $\mathbf{h}$ have periodic support, the fact that $\mathbf{g} * \tilde{\mathbf{h}}$ has aperiodic support adds at least one additional distinct bilinear constraint, because of which the dimension further reduces by at least 1. $\square$

9.3 Support Recovery Proofs

In this section, we assume that V is a k-element subset of $\{0, 1, ..., n-1\}$ chosen uniformly at random, $k = O(n^{1/2-\epsilon})$ and n is large enough.

Lemma 9.4. *Probability that an integer l belongs to W is less than or equal to $\frac{k^2}{n} + o(\frac{k^2}{n})$ if l is not in U.*

Proof. If $l \notin U$, probability that an index l belongs to W can be bounded as follows

$$Pr\{l \in W \ | l \notin U\} = Pr\{\bigcup_g g, g+l \in V | l \notin U\} \leq \frac{k^2}{n} + o\left(\frac{k^2}{n}\right)$$

$\square$

Lemma 9.5. *The probability that an integer l belongs to $W \cap W_1$ is less than or equal to $\frac{k^4}{n^2} + o(\frac{k^4}{n^2})$ if l is not in U.*

Proof.

$$Pr\{l \in W \cap W_1 | l \notin U\} = Pr\{l, l - u_{01} \in W | l \notin U\}$$

$$\leq \sum_d Pr\{u_{01} = d\} \left(\sum_{g,h} Pr\{g, g+l, h, h+l-d \in V | (0,d) \in U, (1, .., d-1, l) \notin U\} \right)$$

$$\leq \sum_d Pr\{u_{01} = d\} \left(\frac{k^4}{n^2} + o\left(\frac{k^4}{n^2}\right) \right) = \frac{k^4}{n^2} + o\left(\frac{k^4}{n^2}\right)$$

□

Lemma 9.6. *The probability that an integer l in W, also belongs to $W \cap W_1$ is less than or equal to $\frac{k^2}{n} + o(\frac{k^2}{n})$ if l is not in U.*

Proof.

$$Pr\{l \in W \cap W_1 | l \in W, l \notin U\} \leq \frac{Pr\{l \in W \cap W_1 | l \notin U\}}{Pr\{l \in W | l \notin U\}} \leq \frac{\frac{k^4}{n^2} + o(\frac{k^4}{n^2})}{\frac{k^2}{n} - o(\frac{k^2}{n})} \leq \frac{k^2}{n} + o\left(\frac{k^2}{n}\right)$$

□

Lemma 9.7. *$Pr\{u_{0t} > u_{k-k^\alpha, k-1}\} \leq \delta_1$ for any $\alpha, \delta_1 > 0$ if $t = log(k)$.*

Proof. Since the integers are chosen uniformly at random, we have

$$E[u_{0t}] = \sum_{i=0}^{i=t-1} E[u_{i,i+1}] \leq \frac{tn}{k} \quad \& \quad var[u_{0t}] = E[u_{0t}^2] - E[u_{0t}]^2 \leq \frac{2tn^2}{k^2} + o\left(\frac{2tn^2}{k^2}\right)$$

Using Chebyshev's inequality, we get

$$Pr\{u_{0t} > \frac{k^{\alpha/2}n}{k}\} \leq \frac{\frac{2tn^2}{k^2} + o\left(\frac{2tn^2}{k^2}\right)}{\left(\frac{k^{\alpha/2}n}{k} - o\left(\frac{k^{\alpha/2}n}{k}\right)\right)^2} \leq \frac{2t}{k^\alpha} + o\left(\frac{2t}{k^\alpha}\right)$$

Similarly, we have

$$E[u_{k-k^\alpha, k-1}] = \frac{(k^\alpha - 1)n}{k} \qquad \& \qquad var\{u_{k-k^\alpha, k-1}\} \leq \frac{2k^\alpha n^2}{k^2} + o\left(\frac{2k^\alpha n^2}{k^2}\right)$$

and

$$Pr\{u_{k-k^\alpha, k-1} < \frac{k^{\alpha/2}n}{k}\} \leq \frac{\frac{2k^\alpha n^2}{k^2} + o\left(\frac{2k^\alpha n^2}{k^2}\right)}{\left(\frac{k^\alpha n}{k} - o\left(\frac{k^\alpha n}{k}\right)\right)^2} \leq \frac{2}{k^\alpha} + o\left(\frac{2}{k^\alpha}\right)$$

Hence, we can bound the desired probability as follows

$$Pr\{u_{0t} > u_{k-k^\alpha, k-1}\} \leq \frac{2t}{k^\alpha} + o\left(\frac{2t}{k^\alpha}\right) < \delta_1$$

for any $\delta_1 > 0$ if n is large enough. Here, we consider the case where k is growing with n, otherwise Corollary 9.1 ensures that the algorithm terminates at Step 2 with the desired solution. □

Lemma 9.8. *The probability that an integer l belongs to $W \cap \left(\bigcap_{p=1}^{t} W_p\right)$ is less than or equal to $(\frac{k^{2(1+\epsilon)}}{n})^{\sqrt{t}/2} + o(\frac{k^{2(1+\epsilon)}}{n})^{\sqrt{t}/2}$ for $t = log(k)$ if l is not in U.*

Proof.

$$Pr\{l \in W \cap \left(\cap_{p=1}^{t} W_p\right) | l \notin U\} = Pr\{l, \{l - u_{0p} : 1 \le p \le t\} \in W | l \notin U\}$$

$$= \sum_{d_1,..d_t} Pr\{u_{0p} = d_p : 1 \le p \le t\} Pr\{l, \{l - d_p : 1 \le p \le t\} \in W | l \notin U, u_{0p} = d_p : 1 \le p \le t\}$$

Using the proofs in [39], we can show that the numbers $\{d_p : 1 \le p \le t\}$ have unique pairwise distances with probability $1 - \delta$ for any $\delta > 0$. Using this property, we see that atleast $\sqrt{t}$ additional vertices in U (apart from $\{u_{0p} : 1 \le p \le t\}$) and atmost $2t$ additional vertices are needed to realize the t pairwise distances. The probability can hence be bounded as

$$\le \sum_{\omega=\sqrt{t}}^{2t} (\omega + t)^t \left(\frac{k}{n}\right)^{\omega} n^{\omega/2} \le (\frac{k^{2(1+\epsilon)}}{n})^{\sqrt{t}/2} + o(\frac{k^{2(1+\epsilon)}}{n})^{\sqrt{t}/2}$$

for $t = log(k)$. □

9.3.1 Proof of Theorem 4.1

The following sequence of Lemmas justify the various steps of Algorithm 2: The fact that u_{01} can be inferred from W is shown in Lemma 9.9. Lemma 9.10 shows that there exist edges between $\{u_{0p} : 1 \le p \le t = log(k)\}$ and $u_{0,k-1}$, and hence $\{u_{0p} : 1 \le p \le t = log(k)\}$ can be inferred from Graph step. The fact that U_t can be calculated using $W \cap \left(\bigcap_{p=1}^{t} W_p\right)$ is proved in Lemma 9.11.

Lemma 9.9 (Step 1)**.** *u_{01} can be inferred from W.*

Proof. The first and second highest terms in W are given by $u_{0,k-1}$ and $u_{1,k-1}$ respectively as $u_{01} \le u_{k-2,k-1}$. Since $u_{01} = u_{0,k-1} - u_{1,k-1}$, we can calculate u_{01} from W. □

Lemma 9.10 (Step 3)**.** *In the graph $G((\{0\} \cup (W \cap W_1)), W)$, integers $\{u_{0p} : 1 \le p \le t = log(k)\}$ have an edge with $u_{0,k-1}$ with probability greater than $1 - \delta$ for any $\delta > 0$.*

Proof. For any fixed p such that $1 \le p \le t$, note that terms u_{0p} and $u_{0,k-1}$ have a pairwise distance $u_{p,k-1}$. For there to be no edge between u_{0p} and $u_{0,k-1}$, another integer pair should have the same pairwise distance. For this to happen, atleast one of the integers should be greater than $u_{p,k-1}$. The only integers greater than $u_{p,k-1}$ in W can be $\{u_{ij} : 0 \le i \le p - 1, j > i\}$. Hence, it is sufficient to show that none of these terms exist in W_1 with the desired probability. These terms can be split into two cases:

(i) $j \le k - k^{\alpha}$:
Using Lemma 9.7, we see that this event can be bounded with probability δ for any $\delta > 0$.

$$u_{ij} \le u_{0t} + u_{pj} < u_{k-k^{\alpha},k-1} + u_{p,k-k^{\alpha}} = u_{p,k-1}$$

(ii) $k - k^\alpha < j \leq k$
There are k^α such terms for a given p. If all the k^α terms don't survive $W \cap W_1$, we can be assured of an edge between u_{0p} and $u_{0,k-1}$.

Hence, if a total of tk^α terms in W don't survive $W \cap W_1$, we are through. The probability that none of them survive $W \cap W_1$ can be upper bounded by $tk^\alpha(\frac{k^2}{n} + o(\frac{k^2}{n}))$ using union bounds and Lemma 9.6. This term can be made less than δ for any $\delta > 0$ if $k = O(n^{1/2-\epsilon})$ and n is large enough. □

Lemma 9.11 (Step 5). *$U_t = W \cap \left(\bigcap_{p=1}^{t} W_p\right)$ with probability greater than $1 - \delta$ for any $\delta > 0$ if $t \geq log(k)$*

Proof. Note that $U_t \subseteq W \cap \left(\bigcap_{p=1}^{t} W_p\right)$ by Intersection step. The extra terms in $W \cap \left(\bigcap_{p=1}^{t} W_p\right)$ can be bounded using the linearity property of expectations and Lemma 9.8 as follows:

$$E[\#l \in W \cap \left(\cap_{p=1}^{t} W_p\right) | l \notin U] \leq n \left(\left(\frac{k^{2(1+\epsilon)}}{n}\right)^{\sqrt{t}/2} + o\left(\frac{k^{2(1+\epsilon)}}{n}\right)^{\sqrt{t}/2} \right) \leq \delta$$

for any $\delta > 0$ if $k = O(n^{1/2-\epsilon})$ and n is large enough. Using Markov inequality, we get

$$Pr\{\left(\#l \in W \cap \left(\cap_{p=1}^{t} W_p\right) | l \notin U\right) \geq 1\} \leq \delta$$

which completes the proof. □

Corollary 9.1. *$U_1 = W \cap W_1$ with probability greater than $1-\delta$ for any $\delta > 0$ if k satisfies $k = O(n^{1/4-\epsilon})$, $n > n(\epsilon, \delta)$.*

Proof. Note that as a consequence, the algorithm can hence be stopped at Step 2 itself if the sparsity $k = O(n^{1/4-\epsilon})$. □

9.4 Signal Recovery Proofs

9.4.1 Proof of Lemma 5.2

Lemma 9.12. *If p denotes the probability that there exists an edge between v_i and v_j in $H(V)$ for $0 \leq i, j \leq k-1$, then $p \geq 1 - \frac{k^2}{n}$.*

Proof. Consider any pair of vertices v_i and v_j. There will be no edge between them if there exists another pair of vertices v_g and v_h such that $v_g - v_h = v_i - v_j$ by construction. Since the support entries are chosen uniformly and randomly, we have

$$Pr\{\exists (g,h) \neq (i,j) \text{ such that } v_g - v_h = v_i - v_j\} \leq \sum_{q=1}^{q=k} \sum_{r=1}^{r=k} Pr\{v_q - v_r = v_i - v_j\} \leq \frac{k^2}{n}$$

$$p = 1 - Pr\{\exists (g,h) \neq (i,j) \text{ such that } v_g - v_h = v_i - v_j\} \geq 1 - \frac{k^2}{n}$$

□

Algorithm 7 Support Recovery: Modified Combinatorial Algorithm

Input: Pairwise distance set W

Output: Integer set U which realizes W

1. Infer u_{01} from W
2. Construct the set $W_1 = W + u_{01}$
3. Construct the graph $G(\{0\} \cup (W \cap W_1))$ and infer $\{u_{0i_p} : 1 \leq p \leq t = log(k)\}$ using Graph step for some $\{i_p : 1 \leq p \leq t = log(k)\}$
4. Construct the set $W_{i_p} = W + u_{0i_p}$ for $1 \leq p \leq t$
5. Calculate $U_{i_t} = W \cap \left(\bigcap_{p=1}^{t} W_{i_p}\right)$
6. Define $\tilde{U} = \{\tilde{u}_0, ...\tilde{u}_{k-1}\}$ as $\tilde{U} = u_{k-1} - U$ and infer $\{\tilde{u}_{0p}$: $1 \leq p \leq t\}$ from U_{i_t}
7. Construct the set $\tilde{W}_p = W + \tilde{u}_{0p}$ for $1 \leq p \leq t = log(k)$
8. Calculate $\tilde{U}_t = \left(\bigcap_{p=0}^{t} \tilde{W}_p\right)$
9. Recover $\tilde{U} = \{\tilde{u}_{0p} : 0 \leq p \leq t-1\} \cup \tilde{U}_t$
10. Recover $U = \tilde{u}_{k-1} - \tilde{U}$

Lemma 9.13. *Suppose $\delta(H(V))$ denotes the minimum degree of the graph $H(V)$, then $\delta(H(V)) \geq k(1 - 1/t)$ with probability $q > 1 - \delta$ for any $\delta > 0$, $t > 0$ and $n > n(\delta)$ if $\frac{k^2 t}{n} < 1$.*

Proof. Consider a vertex v_i. Construct a graph H_i from $H(V)$ by removing all the edges which do not involve the vertex v_i. Let us consider the vertex exposure martingale [49] on this graph H_i with the graph function $d(v_i)$, where $d(v)$ denotes the degree of the vertex v. Let F_j be the induced subgraph of H_i formed by exposed vertices after j exposures. We define a martingale $X_0, X_1, ...X_k$ as follows

$$X_j = E[d(v_i)|F_j]$$

Refer to Section 9.4.3 for details. We have $X_0 = E[d(v_i)] \geq k(1 - \frac{k^2}{n})$ and $X_k = d(v_i)$. Note that $|X_{j+1} - X_j| \leq 1 \quad \forall \quad 0 \leq j \leq m-1$. Azuma's inequality [49] gives us

$$Pr\{d(v_i) < E[d(v_i)] - \lambda\} \leq 2e^{-\lambda^2/2k}$$

for $\lambda > 0$. Choosing $\lambda = k(\frac{1}{t} - \frac{k^2}{n})$, which is greater than 0 when $\frac{k^2 t}{n} < 1$, we get

$$Pr\{d(v_i) < k(1 - \frac{1}{t})\} \leq 2e^{-\frac{k}{2}(\frac{1}{t} - \frac{k^2}{n})^2}$$

Using union bound to accommodate all the vertices v_i for $i = \{0, 1...k-1\}$, we get

$$Pr\{\exists i \text{ such that } d(v_i) < k(1-\frac{1}{t})\} \leq \sum_{i=1}^{i=k} P\{d(v_i) < k(1-\frac{1}{t})\}$$

$$\leq 2ke^{-\frac{k}{2}(\frac{1}{t}-\frac{k^2}{n})^2} < \delta \quad \text{for} \quad n > n(\delta)$$

if $\frac{k^2 t}{n} < 1$. □

Suppose $k = O(n^{1/2-\epsilon})$. Choose $t = 2$ and note that all the conditions of Lemma 9.13 are met, and hence every vertex in the graph has a degree at least $\frac{k}{2}$ with very high probability. Dirac's theorem [50] states that such graphs have a Hamiltonian cycle, which shows that the graph is connected. The probability that there doesn't exist a triangle between any three vertices chosen can be upper bounded by $\frac{3k^2}{n}$ using union bound. Hence, we have

$$q \geq 1 - (2ke^{-\frac{k}{2}(\frac{1}{2}-\frac{k^2}{n})^2} + \frac{3k^2}{n}) > 1 - \delta \quad \text{if} \quad n > n(\delta)$$

9.4.2 Proof of Lemma 5.3

Analysis of semidefinite relaxation based programs with such deterministic measurements is a difficult task in general. We will instead analyze a further relaxation of (17) and show that the relaxed program (20) has a unique optimizer that is rank one and also a feasible point of (17) with the desired probability, which is sufficient to prove Lemma 5.3.

Note that if there exists an edge between vertices v_i and v_j in the graph $H(V)$, then $X_{v_i v_j}$ can be deduced from the autocorrelation. This is because if there is an edge between v_i and v_j, then $a_{|v_i - v_j|} = x_{v_i} x_{v_j}$, which by definition is $X_{v_i v_j}$. A further relaxation of (17) can hence be obtained by using only the values of $\mathbf{X}$ which can be directly inferred from the autocorrelation, i.e,

$$\begin{aligned} &\text{minimize} && trace(\mathbf{X}) \\ &\text{subject to} && X_{v_i v_j} = a_{|v_i - v_j|} \quad \text{if } v_i \leftrightarrow v_j \text{ in } H(V) \\ &&& X_{ij} = 0 \quad \text{if} \quad \{i,j\} \notin V \\ &&& \mathbf{X} \succcurlyeq 0 \end{aligned} \tag{19}$$

where $v_i \leftrightarrow v_j$ means that there exists an edge between v_i and v_j in $H(V)$. Observe that the autocorrelation constraints either fix some components of $\mathbf{X}$ or fix the summation of some components of $\mathbf{X}$, we have discarded the latter to obtain (19). If we further relax (19), by replacing the positive semidefinite constraint with the constraint that every 2×2 submatrix of $\mathbf{X}$ is positive semidefinite, we obtain

$$\begin{aligned} &\text{minimize} && trace(\mathbf{X}) \\ &\text{subject to} && X_{v_i v_j} = a_{|v_i - v_j|} \quad \text{if } v_i \leftrightarrow v_j \text{ in } H(V) \\ &&& X_{ij} = 0 \quad \text{if} \quad \{i,j\} \notin V \\ &&& X_{ii} X_{jj} \geq X_{ij}^2 \quad \forall \text{ distinct } (i,j) \\ &&& X_{ii} \geq 0 \quad \forall\ i \end{aligned} \tag{20}$$

To analyze this program, first consider the following matrix completion problem: let $\mathbf{R}_0 = \mathbf{r}\mathbf{r}^T$ be a positive semidefinite $t \times t$ matrix with all the off-diagonal components known, where $\mathbf{r} = (r_0, ...r_{t-1})$ is a $t \times 1$ vector. The objective is to recover the diagonal components in a robust manner by solving a convex program. Since $\mathbf{R}$ is positive semidefinite, any 2×2 submatrix of $\mathbf{R}$ is also positive semidefinite, i.e.,

$$R_{ii}R_{jj} \geq (r_i r_j)^2 \quad \forall \text{ distinct } (i, j)$$

Consider the convex program

$$\begin{aligned} &\text{minimize} && trace(\mathbf{R}) \\ &\text{subject to} && R_{ii}R_{jj} \geq (r_i r_j)^2 \quad \forall \text{ distinct } (i, j) \\ &&& R_{ii} \geq 0 \quad \forall\, i \end{aligned} \tag{21}$$

Lemma 9.14. *$\mathbf{R}_0$ is the optimizer of (21) with probability greater than $1 - \delta$ for any $\delta > 0$ if t is sufficiently large.*

Proof. Suppose $\mathbf{R}_0 = \mathbf{r}\mathbf{r}^T$ is not the (unique) minimizer of (21). Then there exists atleast one i such that $R^\star_{ii} < r_i^2$, say $R^\star_{ii} = (1-\gamma)r_i^2$ for some $\gamma > 0$. The constraints require all other diagonal components to be greater than or equal to $\frac{1}{1-\gamma} r_j^2$, which in turn is greater than $(1+\gamma)r_j^2$. The objective function value at the optimum can be written as

$$\begin{aligned} trace(\mathbf{R}^\star) &= \sum_{i=1}^{i=t} R^\star_{ii} > (1-\gamma)r_i^2 + \sum_{j \neq i}(1+\gamma)r_j^2 \\ &= \sum_j {r_j}^2 + \gamma\Big(\sum_{j \neq i} r_j^2 - r_i^2\Big) \end{aligned}$$

If we can ensure that $(\sum_{j \neq i} r_j^2 - r_i^2) > 0$ for all i, we are through because $trace(\mathbf{R}^\star)$ is greater than $\sum_j r_j^2$, which is a contradiction. [52] provides an exponentially decreasing probability in t for failure of this condition. □

Lemma 9.13 shows that $\delta(H(V)) > k(1 - \frac{1}{t})$ if $\frac{k^2 t}{n} < 1$. Hajnal-Szemeredi theorem on disjoint cliques [51] states that such graphs contain $\frac{k}{t}$ vertex disjoint union of complete graphs of size t. Suppose we choose $t = log(n)$. Lemma 9.14 applies to each of the $\frac{k}{t}$ complete graphs and hence using union bound, we see that the optimizer of (20) has all the diagonal components corresponding to the rank 1 completion with very high probability for $k = O(n^{1/2-\epsilon})$. Since the graph $H(V)$ is connected, we know both the diagonal and the principal off-diagonal components of the optimizer. Since they have components from a rank 1 matrix, the rank 1 completion is the only positive semidefinite completion, and hence the unique minimizer of (20). Since the optimizer also satisfies all the constraints of (17), it is the unique minimizer of (17).

9.4.3 Vertex exposure martingale

We can see that $X_0, X_1, ...X_k$ form a martingale as follows

$$E[X_{j+1}|X_0, X_1, ...X_j] = E[E[d(v_i)|H_{j+1}]|X_0, X_1, ...X_j]$$

$$= \sum_{H_{j+1}} p(H_{j+1}|X_0, X_1, ...X_j)E[d(v_i)|H_{j+1}] = \sum_{H_{j+1}} p(H_{j+1}|H_j) \sum_{G'} p(G'|H_{j+1})d_{G'}(v_i)$$

where $d_{G'}(v_i)$ is the degree of v_i in the graph G' and the summation is done over all possible graphs G'.

$$= \sum_{G'} \sum_{H_{j+1}} p(G', H_{j+1}|H_j)d_{G'}(v_i) = E[d(v_i)|H_j] = X_j$$

9.5 Noise Robustness Proofs

9.5.1 Proof of Theorem 6.1

Suppose the measurements have additive noise $\mathbf{n}$ such that $||\mathbf{n}||_\infty < \eta$, where η is less than half of the minimum non-zero entry of the autocorrelation in magnitude. The set W in this case can be calculated exactly by using the indices of the elements in the autocorrelation which have a value greater than η in magnitude. In order to analyze the error in the recovered signal values, we use a technique similar to the proof of Lemma 5.3. We analyze (22), which is a relaxation of (18).

$$\begin{aligned} &\text{minimize} && trace(\mathbf{X}) \\ &\text{subject to} && ||X_{ij} - a_{|u_i-u_j|}|| < \eta \quad \text{if } u_i \leftrightarrow u_j \text{ in } H(U) \\ &&& X_{ij} = 0 \quad \text{if} \quad \{i,j\} \notin U \\ &&& \mathbf{X} \succcurlyeq 0 \end{aligned} \tag{22}$$

Consider the following matrix completion problem: let $\mathbf{R}_0 = \mathbf{r}\mathbf{r}^T$ be a $t \times t$ matrix whose off-diagonal components are measured with additive noise, i.e., $Q_{ij} = R_{ij} + n_{ij} \quad \forall$ distinct (i,j), where $\mathbf{r} = (r_0, ...r_{t-1})$ is a $t \times 1$ vector and the noise $\mathbf{N}$ satisfies $||\mathbf{N}||_\infty < \eta < |r_i|^2/2 \quad \forall i$. The objective is to recover the diagonal components in a robust manner by solving a convex program. Consider the convex program

$$\begin{aligned} &\text{minimize} && trace(\mathbf{R}) \\ &\text{subject to} && ||Q_{ij} - R_{ij}|| < \eta \\ &&& \mathbf{R} \succcurlyeq 0 \end{aligned} \tag{23}$$

For the optimizer of (23), we have

$$|R^\star_{ij} - r_i r_j| \le |R^\star_{ij} - Q_{ij}| + |Q_{ij} - r_i r_j| < 2\eta \quad \forall \text{ distinct } (i,j)$$

Since $\eta < |r_i|^2/2 \quad \forall i$, we have the bound $(R^\star_{ij})^2 \ge (r_i r_j - 2\eta)^2 \ge (r_i^2 - 2\eta)(r_j^2 - 2\eta)$.

Lemma 9.15. *If $\mathbf{R}^\star$ is the optimizer of (23), then $\sum_j |R^\star_{jj} - R_{0,jj}| \le 4t\eta$ with arbitrarily high probability for sufficiently large t.*

Proof. Let $\mathbf{R}^\star$ be the optimal solution to (23). Since for all off-diagonal components, we have $(R^\star_{ij})^2 \ge (r_i^2 - 2\eta)(r_j^2 - 2\eta)$, there cannot exist i and j such that $R^\star_{ii} < r_i^2 - 2\eta$ and $R^\star_{jj} < r_j^2 - 2\eta$ due to the fact that 2×2 submatrices of $\mathbf{R}^\star$ have to be positive semidefinite. If one of the diagonal terms (say i) is such that $R^\star_{ii} = (1-\gamma)(r_i^2 - 2\eta)$, then the 2×2 positive semidefinite constraints would ensure that for all $j \neq i$, $R^\star_{jj} > (1+\gamma)(r_j^2 - 2\eta)$. The optimum value would hence be

$$trace(\mathbf{R}^\star) > (1-\gamma)(r_i^2 - 2\eta) + \sum_{j \neq i}(1+\gamma)(r_j^2 - 2\eta) = \sum_j r_j^2 + \gamma(\sum_{j \neq i} r_j^2 - r_i^2) - 2t\eta - 2(t-2)\gamma\eta$$

which, similar to the proof of Lemma 5.3, is strictly greater than $\sum_j r_j^2$ with arbitrarily high probability, which is a contradiction. Hence, the optimizer has diagonal components $R^\star_{jj} \geq r_j^2 - 2\eta \quad \forall j$ with arbitrarily high probability. Since the objective function value at the optimizer is less than or equal to $\sum_j r_j^2$ (because $\mathbf{R}_0$ is a feasible point with objective function value $\sum_j r_j^2$), $\sum_j |R^\star_{jj} - R_{0,jj}| \leq 4t\eta$ $\square$

Using the same arguments as Lemma 5.3, we can see that the optimizer of (22) and (18) satisfy $\sum_j |X^\star_{jj} - X_{0,jj}| \leq 4k\eta$, or in other words, $|||\hat{\mathbf{x}}| - |\mathbf{x}|||_2^2 \leq 4k\eta$. Note that the diagonal components only ensure that the magnitude of the signal values can be recovered robustly. The bound on η ensures that noise does not alter the signs of the measured terms in the matrix, hence the relative signs of the various components can be recovered accurately, because of which $||\hat{\mathbf{x}} - \mathbf{x}||_2^2 \leq 4k\eta$.

References

[1] A. L. Patterson: A direct method for the determination of the components of interatomic distances in crystals, Zeitschr. Krist. 90 (1935) 517-542

[2] A.L. Patterson: Ambiguities in the X-ray analysis of crystal structures, Phys. Review 65 (1944) 195-201.

[3] R. P. Millane, "Phase retrieval in crystallography and optics," J. Opt. Soc. Am. A 7, 394-411 (1990)

[4] J.C. Dainty and J.R. Fienup,‘ ‘Phase Retrieval and Image Reconstruction for Astronomy," Chapter 7 in H. Stark, ed., Image Recovery: Theory and Application pp. 231-275.

[5] L. Rabiner and B.H. Juang, “Fundamentals of Speech Recognition,” Signal Processing Series, Prentice Hall, 1993.

[6] A. Walther, "The question of phase retrieval in optics," Opt. Acta 10, 4149 (1963).

[7] M. Stefik, “Inferring DNA structures from segmentation data”, Artificial Intelligence 11 (1978).

[8] R. W. Gerchberg and W. O. Saxton. “A practical algorithm for the determination of the phase from image and diffraction plane pictures”. Optik 35, 237 (1972).

[9] J. R. Fienup, “Phase retrieval algorithms: a comparison”. Appl. Opt. 21, 2758–2769 (1982).

[10] H.H. Bauschke, P.L. Combettes and D.R. Luke, "Phase retrieval, error reduction algorithm, and Fienup variants: a view from convex optimization" J. Opt. Soc. Am.A (2002).

[11] Y.M. Lu and M. Vetterli. “Sparse spectral factorization: Unicity and reconstruction algorithms”. ICASSP 2011.

[12] E. J. Candes, Y. Eldar, T. Strohmer and V. Voroninski, "Phase retrieval via matrix completion", arXiv:1109.0573 [cs.IT], 2011.

[13] Y. Shechtman, Y.C. Eldar, A. Szameit and M. Segev, "Sparsity Based Sub-Wavelength Imaging with Partially Incoherent Light Via Quadratic Compressed Sensing", Optics Express, vol. 19, Issue 16, pp. 14807-14822, Aug. 2011.

[14] K. Jaganathan, S. Oymak and B. Hassibi, "Recovery of Sparse 1-D Signals from the Magnitudes of their Fourier Transform". ISIT 2012.

[15] E. J. Candes, T. Strohmer, and V. Voroninski, Phase lift: Exact and stable signal recovery from magnitude measurements via convex programming, arXiv:1109.4499, Sept. 2011.

[16] Y. Shechtman, A. Beck and Y. C. Eldar, "GESPAR: Efficient Phase Retrieval of Sparse Signals", arXiv:1301.1018 [cs.IT].

[17] Y.C. Eldar and S. Mendelson, "Phase retrieval: Stability and recovery guarantees," arXiv preprint arXiv:1211.0872 (2012).

[18] P. Netrapalli, P. Jain and S. Sanghavi, "Phase Retrieval using Alternating Minimization," arXiv preprint arXiv:1306.0160 (2013).

[19] H. Ohlsson, A. Yang, R. Dong, and S. Sastry, "Compressive phase retrieval from squared output measurements via semidefinite programming, arXiv preprint arXiv:1111.6323, 2011.

[20] P.Schniter and S.Rangan,Compressive phase retrieval via generalized approximate message passing, in 50th Annual Allerton Conference on Communication, Control, and Computing, 2012.

[21] A. Fannjiang. "Absolute Uniqueness of Phase Retrieval with Random Illumination". arXiv:1110.5097v3

[22] S. Mukherjee and C. Seelamantula, "An iterative algorithm for phase retrieval with sparsity constraints: Application to frequency domain optical coherence tomography, in Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on. IEEE, 2012, pp. 553556.

[23] E. Candes, X. Li, and M. Soltanolkotabi. "Phase Retrieval from masked Fourier transforms." arXiv preprint arXiv:1310.3240 (2013).

[24] S. Oymak, A. Jalali, M. Fazel, Y.C. Eldar, and B. Hassibi, "Simultaneously Structured Models with Application to Sparse and Low-rank Matrices", preprint available at arXiv:1212.3753.

[25] X. Li, V. Voroninski, " Sparse Signal Recovery from Quadratic Measurements via Convex Programming", arXiv:1209.4785v1.

[26] A. S. Bandeira, Y. Chen and D. G. Mixon, "Phase retrieval from power spectra of masked signals", arXiv:1303.4458 [math.FA]

[27] M. Hayes and J. McClellan, "Reducible Polynomials in more than One Variable", Proc. IEEE 70(2): 197 198, (1982)

[28] J. Ranieri, A. Chebira, Y. M. Lu and M. Vetterli, "Phase Retrieval for Sparse Signals: Uniqueness Conditions." arXiv preprint arXiv:1308.3058 (2013).

[29] H. Ohlsson and Y. C. Eldar, "On Conditions for Uniqueness in Sparse Phase Retrieval", arXiv:1308.5447 [cs.IT].

[30] A.H. Sayed and T. Kailath. "A survey of spectral factorization methods". Numerical Linear Algebra with Applications, 8: 467–496 (2001).

[31] B. Recht, M. Fazel, and P. Parrilo. "Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization". SIAM Review, Vol 52, no 3.

[32] M. A. Khajehnejad, W. Xu, S. Avestimehr and B. Hassibi, "Weighted l_1 Minimization for Sparse Recovery with Prior Information", arXiv: 0901.2912 [cs.IT]

[33] E. J. Candes, M. B. Wakin, S. P. Boyd, "Enhancing Sparsity by Reweighted l1 Minimization", Journal of Fourier Analysis and Applications (2008).

[34] M. Fazel, H. Hindi, and S. Boyd. "Log-det Heuristic for Matrix Rank Minimization with Applications to Hankel and Euclidean Distance Matrices". Proc. American Control Conference, 2003.

[35] E. J. Candes and T. Tao. "Decoding by linear programming". IEEE Trans. Inform. Theory, 51 4203–4215.

[36] H. Ohlsson, A. Yang, R. Dong, M. Verhaegen and S. Sastry, "Quadratic Basis Pursuit, arXiv:1301.7002 [cs.IT].

[37] K. Jaganathan, S. Oymak and B. Hassibi, "Sparse Phase Retrieval: Convex Algorithms and Limitations", arXiv:1303.4128 [cs.IT].

[38] J. R. Fienup, T. R. Crimmins, and W. Holsztynski. "Reconstruction of the support of an object from the support of its autocorrelation", J. Opt. Soc. Am. 72, 610–624 (1982)

[39] K.Jaganathan, S. Oymak and B.Hassibi, "On Robust Phase Retrieval for Sparse Signals", Allerton 2012.

[40] S.S. Skiena, W.D. Smith and P. Lemke, "Reconstructing Sets from Interpoint Distances (Extended Abstract)", SCG' 90 Proceedings of the sixth annual symposium on computational geometry, Pages 332-339.

[41] A. Ahmed, B. Recht and J. Romberg, "Blind Deconvolution using Convex Programming", arXiv:1211.5608 [cs.IT].

[42] J. Rosenblatt, P.D. SEYMOUR, "The structure of Homometric Sets", SIAM J. Alg. Disc. Meth, Vol 3, No 3, September 1993.

[43] P . Lemke ,M. Wermano, "On the complexity of inverting the autocorrelation function of a finite integer sequence, and the problem of locating n points on a line, given the unlabeled distances between them", unpublished manuscript (1988)

[44] Z. Zhang , "An exponential exarnple for partial digest mapping algorithm", Journal of Cornputational Biology 3 (1994), 235239

[45] T. Dakic, "On the Turnpike Problem", PhD Thesis, Simon Fraser University, 2000.

[46] R.M. Karp and L.A. Newberg, " An Algorithm for Analyzing Probed Partial Digestion Experiments", Comput Appl Biosci (1995) 11(3): 229-235 doi:10.1093/bioinformatics/11.3.229.

[47] G. Pandurangan and H.Ramesh, "The Restriction Mapping Problem Revisited", Journal of Computer and System Sciences, 2002.

[48] K. Jaganathan and B. Hassibi, "Reconstruction of Integers from Pairwise Distances", arXiv:1212.2386 [cs.DM].

[49] N. Alon and J.H. Spencer, The Probabilistic Method, Chapter 7.

[50] R.L. Graham, Handbook of Combinatorics

[51] Hajnal, A. and Szemeredi, E. Proof of a Conjecture of Erdos. In Combinatorial Theory and Its Applications, Vol. 2 (Ed. P.Erdos, A.Renyi and V.T. Sos). Amsterdam, Netherlands: North-Holland, 1970.

[52] R. Vershynin, Introduction to the non-asymptotic analysis of random matrices. arXiv: 1011.3027v7.